

The Catalan butterfly monitoring scheme has the capacity to detect effects of modifying agricultural practices

MARINA S. LEE^{1,†}, JORDI COMAS², CONSTANTÍ STEFANESCU³, AND RAMON ALBAJES¹

¹AGROTECNIO Center, Universitat de Lleida, Rovira Roure 191, 25198 Lleida, Spain

²Departament d'Enginyeria Agroalimentària i Biotecnologia, Universitat Politècnica de Catalunya, Esteve Terrades, 8, 08860 Castelldefels, Barcelona, Spain

³Butterfly Monitoring Scheme, Museu Granollers-Ciències Naturals, Francesc Macià, 51, ES-08402 Granollers, Spain

Citation: Lee, M. S., J. Comas, C. Stefanescu, and R. Albajes. 2020. The Catalan butterfly monitoring scheme has the capacity to detect effects of modifying agricultural practices. *Ecosphere* 11(1):e03004. 10.1002/ecs2.3004

Abstract. Impacts of agricultural management practices on the receiving environment are seldom suitably assessed because environmental monitoring is costly. In this regard, data generated by already existing environmental survey networks (ESNs) may have sufficient capacity to detect effects. Here, we study the capacity of the Catalan butterfly monitoring scheme (CBMS) to detect differences in butterfly abundance due to changes in agricultural practices. As a model, we compared butterfly abundance across two landscape types according to agricultural intensification. A 2 km diameter buffer area was centered on the CBMS transect, the control group were transects located in areas where intensive agriculture represented <20% of the area; a treated group was simulated by selecting transects located in areas where intensive agriculture occupied an area over 40%. The Welch *t*-test ($\alpha = 0.05$ and 80% power) was used to compare butterfly abundance per section across landscape types. The capacity of the *t*-test to detect changes in mean butterfly abundance, of 12 butterfly indicators relevant to farmland, was calculated annually and for 5-, 10-, and 15-yr periods. Detection capacity of the *t*-test depended mainly on butterfly data sample size and variability; difference in butterfly abundance was less important. The *t*-test would be capable of detecting acceptably small population changes across years and sites. For instance, considering a 15-yr period, it would be possible to detect a change in abundance below 10% of the multispecies indicators (all butterfly species, open habitat species, mobile species, and grassland indicators) and two single species (*Lasiommata megera* and *Lycaena phlaeas*). When comparisons were carried out within each year, the *t*-test would only be capable of detecting a change below 30% for all butterfly species, mobile species, and *L. megera*. However, detection capacity rapidly improved with the addition of further years, and with 5 yr of monitoring, all indicators but *Thymelicus acteon* had a detection capacity below 30%. We therefore conclude that, from a statistical point of view, the CBMS data “as is” are sensitive enough for monitoring effects of changes in agricultural practices. It could be used, for instance, for the general surveillance of genetically modified crops.

Key words: agriculture; butterfly; detection capacity; environmental monitoring; general surveillance; GM maize; impact assessment; *Lasiommata megera*; *Lycaena phlaeas*; power; *t*; test.

Received 23 October 2019; **accepted** 28 October 2019. Corresponding Editor: Debra P. C. Peters.

Copyright: © 2020 The Authors. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

† **E-mail:** marinaleecrombie@yahoo.es

INTRODUCTION

Changes in agricultural management practices can affect the capacity of the receiving environment to deliver ecosystem services that are essential to maintain the productivity of agricultural land (i.e., Tschamntke et al. 2005). Therefore, in order to manage our resources appropriately (Ripple et al. 2017), it is essential to monitor indicators capable of providing reliable information on the state of the environment before and after changes in agricultural practices (Elzinga et al. 2001). However, environmental monitoring is often too costly to implement because a high number of replications in space and time are needed for reliably detecting changes of the indicators (Field et al. 2007).

For this reason, there is considerable interest in using data collected by already existing environmental survey networks (ESNs) for environmental impact assessment. The use of ESNs has received much attention (Morecroft et al. 2009, Geijzendorffer and Roche 2013), for instance, recently for monitoring the impact of GMOs on natural communities (Lang and Bühler 2012, EFSA 2014). However, their applicability for monitoring is still uncertain (i.e., Smets et al. 2014), particularly regarding the suitability of ESN data for statistical analysis of effect detection capacity.

In order to use ESNs' data for detecting impacts, measurable effects on the receiving environment must be detected (Field et al. 2007). After setting the degree of change (effect size) considered sufficient to trigger a management response, the most fundamental requirement is that the data should be capable of detecting the change if it actually occurs, that is, that it will yield adequate statistical power. For certain impacts of agricultural practices like GMOs on nontarget organisms, capacity to detect population changes of 25–50% has been considered acceptable in field trials (Duan et al. 2006, Lopez et al. 2005, Perry et al. 2003). Unfortunately, most ESNs lack sufficient statistical power to prevent false-negative conclusions when temporal (e.g., before and after implementation of novel crop management practices) and spatial (areas where the measure has been introduced compared to areas where it has not) average differences are compared (Hails et al. 2012).

Here, we carry out a case study using data from the Catalan butterfly monitoring scheme (henceforth CBMS), a large-scale network based on transect counts, to determine its capacity to detect changes in mean abundance of butterfly populations due to changes in agricultural practices. This network is particularly relevant because it uses a standardized monitoring protocol for collection of butterfly data (Van Swaay et al. 2008); butterflies are widely recognized ecological indicators, capable of reflecting environmental impacts of human activity in terrestrial ecosystems (Thomas 2005); and the CBMS is located in a region where high biodiversity and agricultural intensification converge.

The main aim of this work was to determine to what extent the CBMS sampling protocol can be used as is for detecting a potential change, from a statistical point of view. In particular, we calculated the capacity of Welch's *t*-test (Welch 1947) to detect eventual changes in butterfly abundance related to agricultural management.

MATERIALS AND METHODS

Butterfly data were provided by the CBMS (www.cbms.org), a network monitoring butterfly populations in Catalonia (NE Iberian Peninsula) since 1994. Butterfly data were standardized, and a two means unequal variance *t*-test was used to compare abundance of four multispecies indicators and eight species, across two broad landscape types. The sensitivity of the *t*-test to detect changes in abundance between samples was calculated for two types of analysis, annual and multiannual (5, 10, and 15 yr of monitoring data).

Butterfly dataset

Butterfly data were provided by the CBMS (CBMS 2016), which currently has over 150 recording transects located throughout Catalonia, Andorra, and the Balearic Islands. The CBMS uses a standardized methodology for data collection (Pollard and Yates 1993), common to most European butterfly monitoring schemes (Schmucki et al. 2016). In short, trained observers count the number of butterflies observed within a 5 × 5 m virtual area along a line transect of approximately 1.5 km, which is divided into

sections of variable length according to the surrounding habitat type. Counts take place weekly from March to September, only in good weather conditions (about 30 counts per year). Transect counts yield species-specific relative abundance indices that are assumed to reflect year-to-year population changes over the entire study area (Pollard and Yates 1993).

For this study, we selected a subset of the CBMS transects (Fig. 1), which were separated into two broad classes, agricultural (Ag) and non-agricultural (Appendix S1) (non-Ag), according to land cover of surrounding landscape, to test for an expected effect of land cover on butterfly communities. We would expect changes in agricultural practices to have a

reduced impact on butterfly populations in non-Ag landscapes. Transects located in urban (more of 20% of the area covered by buildings and associated infrastructure) and montane (above 800 m) areas were excluded. Transects were also required to have been operative for at least 10 yr (data were analyzed from 1999 to 2013). This resulted in 11 Ag transects located in agricultural landscapes (i.e., where the area of arable crops and orchards was above 40% of a total area determined by a 2-km circle centered on the transect) and 18 non-Ag transects that were located in areas where arable crops and orchards accounted for <20%; non-Ag landscapes were dominated by grassland, scrub, or forest and were often within protected areas. Land-use

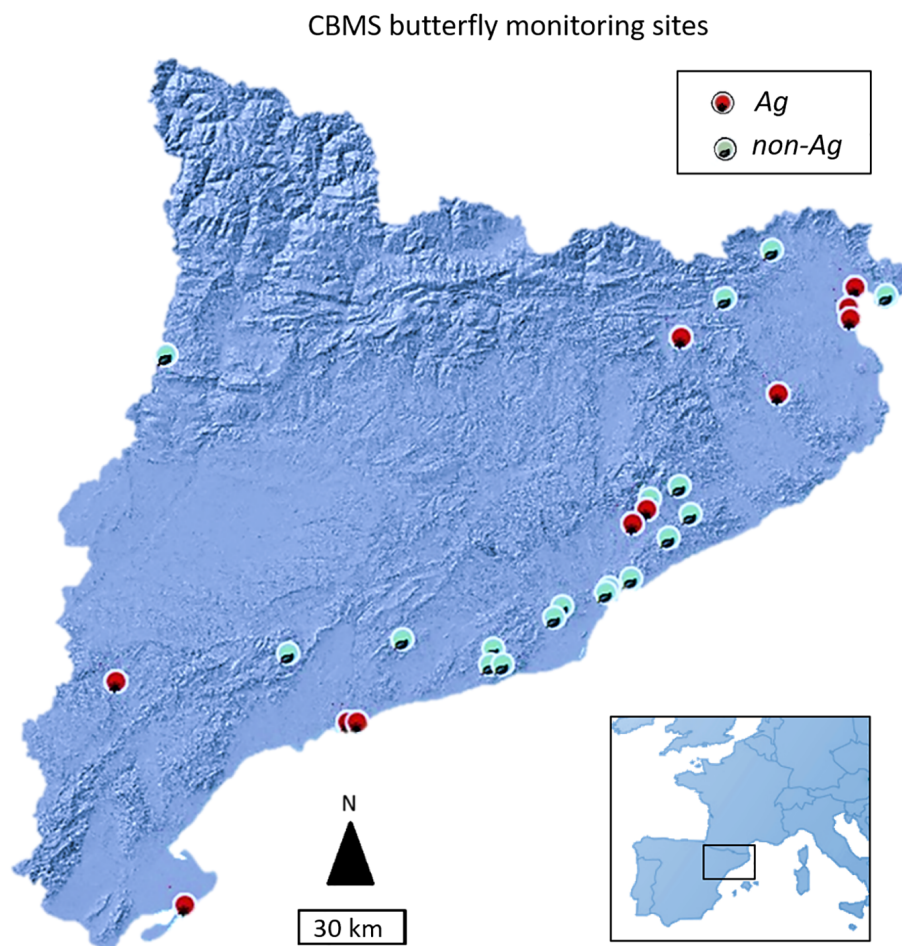


Fig. 1. Location of the 29 Catalan butterfly monitoring scheme (CBMS) transects analyzed, grouped according to the intensity of agriculture in the surrounding landscape into agricultural (Ag) transects and non-agricultural (non-Ag) transects.

cover was mapped and calculated using ARCGIS version 9.3 (ESRI 2008) on the basis of georeferenced aerial photographs (ICGC 2013). For each butterfly indicator, only sections with nonzero values were used for the analysis (i.e., Lang 2004) as the focus of this study was on abundance data, rather than presence–absence of species (Elzinga et al. 2001).

The dataset analyzed comprised 135 butterfly species, many of which were present in only a few transects or years. This resulted in frequent gaps in the data making it impossible to compare populations of specific species across diverse geographical areas. To address this problem, four multispecies indicators were generated by aggregating species according to ecological traits relevant for monitoring agricultural impacts (see Supporting Information in Appendix S2 for multispecies indicator composition). Multispecies indicators “all species” included all butterfly counts (135 species); “open habitat species” comprised 41 butterfly species associated with open habitats (Herrando et al. 2016) after excluding 12 species with a documented strong migratory behavior, whose abundance depends heavily on the conditions at their place of origin (Stefanescu et al. 2011b); “mobile species” included 23 species with a high dispersal ability (Stefanescu et al. 2011a); and finally, “grassland indicators” aggregated 16 species from the European Grassland Indicator, developed by the European Environment Agency (EEA 2013). However, although monitoring multispecies indicators can produce very good results from a statistical point of view, the interpretation of results is more straightforward when monitoring single species. For this reason, single species potentially suitable for monitoring impacts in agricultural land were selected from the 15-yr dataset. Candidate species were required to have a relative detection capacity of Welch’s *t*-test below 30% (see following sections), migrants were excluded, and we selected eight single species that were widespread and common in farmland in the study area (Lee and Albajes 2013); butterfly nomenclature followed Van Swaay et al. (2010).

Data analysis

Before the analysis, butterfly abundance was standardized to density (individuals/km) by

dividing the sections’ butterfly abundance by section length and multiplying by 1000. Data were not transformed because for large datasets, the calculated means and their deviations approximate to the normal distribution (Haze-winkel 2002). After standardization, mean butterfly abundance per transect section (counts from March to September) was calculated for each year (1999–2013). Only nonzero transect counts were used for calculations because this study focuses on the detection of change by comparing abundance only across sections where the species is present. The drawback of this method is that local extinction of species could be overlooked; nevertheless, the detection of a population decline can be used as an early warning. Standardized mean annual abundance of each butterfly indicator per section was the basic unit used in both annual and multiannual analyses.

Contrast statistic.—To simulate a situation in which differences would exist between Ag (potentially disturbed by modified agricultural practices) and non-Ag landscapes (less likely exposed to disturbances by modified agricultural practices), we tested the hypothesis that there were no differences between means of sections in the two landscape types. For each butterfly indicator, the mean abundance per section in each landscape type was compared within each year (annual analysis) and also aggregating means from the 5-, 10-, and the entire 15-yr period (multiannual analysis). We used a two-sided *t*-test without assuming equal variances (Welch 1947), a robust technique for large datasets (Fagerland 2012), which is commonly used for field testing effects of agricultural practices (i.e., Feber et al. 2007, Aviron et al. 2009, Lang and Bühler 2012). The contrast statistic (*t*) was calculated by

$$t = \frac{(m_{\text{Ag}} - m_{\text{non-Ag}})}{\sqrt{\frac{s_{\text{Ag}}^2}{n_{\text{Ag}}} + \frac{s_{\text{non-Ag}}^2}{n_{\text{non-Ag}}}}} \quad (1)$$

where m_{Ag} , s_{Ag} , and n_{Ag} ; and $m_{\text{non-Ag}}$, $s_{\text{non-Ag}}$, and $n_{\text{non-Ag}}$ are the mean, standard deviation, and sample size for samples in Ag and non-Ag landscapes, respectively. The denominator is the standard error (SE) of the statistic $m_{\text{Ag}} - m_{\text{non-Ag}}$. The significance of the test is expressed by the *P*-

value ($\alpha = 0.05$). Correlations between longitudinal series (across years) and horizontal sampling data (between sections within a transect) were checked for lack of temporal and spatial autocorrelation.

Detection capacity of the t-test.—Following Albaladejo et al. (2013), the detection capacity of the *t*-test was computed by establishing *ex ante* the probability of false positives (i.e., the probability of the test producing a significant result when there are no differences between the means of the two populations, symbolized by α) and false negatives (i.e., the probability of the test not producing a significant result, when population means are not equal, symbolized by β).

The detection capacity (*D*) of Welch's *t*-test expressed in absolute terms was computed as follows:

$$D = (t_{(1-\alpha)/2} - t_b) \times \sqrt{\frac{s_{Ag}^2}{n_{Ag}} + \frac{s_{non-Ag}^2}{n_{non-Ag}}} \quad (2)$$

where variables were defined as in Eq. 1; α was set at 0.05 and β at 0.2, values considered acceptable in field tests (Perry et al. 2003), but that can be modified if required (Di Stefano 2003, Field et al. 2007). According to this procedure, the detection capacity is the size of the population change (effect size) of a given species or group that could be detected given its abundance, variability, and sample size. This expression may also be used to calculate the relative detection capacity of the test in relation to the mean abundance in the control (non-Ag landscapes), D_N . Further details regarding this procedure can be found in Comas et al. (2013).

All calculations and statistical analysis were done using the R software (R Core Team 2016); *t*-tests were carried out with R Stats Package version 3.3.1 and detection capacity calculations based on Package pwr version 1.2-1.

RESULTS

The mean length of the 29 transects selected was 1692 ± 132 m (mean \pm SD); each transect was divided into 5–16 sections (mean length 198 ± 63 m). The raw dataset consisted of 262,044 butterfly section counts, 102,210 from the

11 Ag transects, and 159,834 from the 18 non-Ag transects. This is the result of considering, for each of the 135 species recorded, the number of transect sections with nonzero values in each sampling date (~30 dates per year), and the number of years (15). In order to compare butterfly abundance across the two landscape types, the annual mean was calculated for each section.

In the 15-yr dataset, 52 species of the 135 had a relative detection capacity (D_N) below 30%; their frequency in section counts was above 34.6% (mean sample size of 504 sections), and SE values were generally below 1.01. From these candidates, eight indicator species widespread in farmland were selected for the annual and multi-annual analyses.

Mean annual sample sizes of the multispecies indicators, all species (135 species), open habitat species (41 species), grassland indicators (16 species), or mobile species (23 species), were roughly similar in both landscape types (Table 1). When the eight single species were analyzed, sample sizes were considerably lower (see Table 1 for mean annual values, the detailed year-by-year analysis is shown in Appendix S3).

When section counts were compared across the 5-, 10-, or the entire 15-yr period (multianual analysis), sample sizes of the butterfly indicators increased greatly (Table 2) in comparison with the annual analysis (Table 1).

Factors that affected the outcome of the t-test

Once the significance and the power of the test have been set, the *t*-test is influenced by the difference between mean abundance of the two samples and the standard error of the test. The SE of Welch's *t*-statistic is the square root of the sum of the ratios between the variance and the size of each sample (Eq. 1). Consequently, the larger the sample size, the smaller the SE of the test. As expected, as the average size of two samples increased, the SE of the contrast statistic decreased exponentially ($R^2 = 0.60$):

$$SE = 10.5 \times n^{-0.45} \quad (3)$$

However, for similar sample sizes, the range of the test's SE values was quite wide, especially when considering the annual analysis (see Table 1; Appendix S3). In multispecies data, SE ranged between 1.01 and 1.83, whereas average

Table 1. Relative detection capacity (D_N) of Welch's t -test ($\alpha = 0.05$, $\beta = 0.2$) when mean butterfly abundance was compared year by year (from 1999 to 2013) across two broad landscape types.

Multispecies groups	Mean sample size (n sections)		Mean abundance (butterflies/km)		Mean SE	Years with significant differences	Mean D_N (%)
	Ag	Non-Ag	Ag	Non-Ag			
All species	91	115	11.3	14.8	1.28	13	24.1
Open habitat species	86	114	12.5	16.9	1.83	9	30.2
Mobile species	88	114	10.0	10.9	1.01	1	26.0
Grassland indicators	85	114	11.4	13.1	1.40	1	30.2
Single species							
<i>Aricia cramera</i>	20	40	9.5	11.1	2.32	3	59.9
<i>Carcharodus alceae</i>	28	27	6.7	10.3	2.49	3	68.1
<i>Lasiommata megera</i>	61	103	8.3	11.7	1.09	15	26.5
<i>Lycaena phlaeas</i>	42	55	7.7	11.0	1.41	11	36.2
<i>Pararge aegeria</i>	62	73	10.3	11.3	1.63	1	40.7
<i>Pieris napi</i>	46	22	12.6	9.6	1.95	5	58.3
<i>Polyommatus icarus</i>	68	76	13.9	16.6	2.92	2	47.6
<i>Thymelicus acteon</i>	27	45	8.9	13.0	3.49	0	71.6

Notes: Relative detection capacity was calculated according to Eq. 2. Ag: sections located in landscapes with prevalence of agricultural habitats; non-Ag: ibid. but non-agricultural habitats.

sample sizes varied in a much narrower range, 100–103, and the difference between mean abundance from 1.70 to 4.36 butterflies per km. Likewise, regarding single species, the highest SE values did not strictly correspond to the lowest sample sizes (Table 1).

In the multiannual analysis (Table 2), similar trends were observed. For instance, considering the entire 15-yr dataset, the values of the test's s_e differed between the four multispecies indicators (0.34, 0.49, 0.27, and 0.36 individuals per km) in spite of their relatively similar sample size (from 1489 to 1540). Examining single species, SE values were again not directly related to sample sizes. The two species with the highest SE (0.82 and 0.92 butterflies per km) had differing sample sizes (average 1086 and 540); similarly, the lowest SE values (0.29 and 0.36 butterflies per km) also corresponded to species with relatively different sample sizes (average 1230 and 730) (Table 2).

Detection capacity of the t -test

We calculated the detection capacities of Welch's t -test ($\alpha = 0.05$ and $\beta = 0.2$), expressed relative to the average abundance in non-Ag landscapes (D_N), of the four multispecies indicators and the eight single species considered. The t -test was carried out firstly comparing annual abundance in non-Ag and Ag landscapes year by

year (Table 1 and Fig. 2A–C, detailed results in Appendix S3), and then, means were compared across an increasing timespan of 5, 10, and 15 yr (Table 2 and Fig. 2D–F).

When comparisons were conducted within each year, differences between the two landscape types were detected only in some cases (Table 1). For the four multispecies indicators, mean annual D_N values were 24.1%, 30.2%, 26.0%, and 30.2% for all species, open habitat species, mobile species, and grassland indicators, respectively; D_N fluctuated always below 31% (Fig. 2A) being lowest for most years for all species and then for mobile species; and the worst D_N values were those of open habitat species and grassland indicators. Regarding the eight single species, average D_N values (Fig. 2B, C) ranged from 26.5% (*Lasiommata megera*) to 71.6% (*Thymelicus acteon*); only four species showed D_N values under 50% (Fig. 2B) in most years, whereas the remaining four species showed very poor D_N values, consistently above 50% (Fig. 2C).

When the mean comparison was carried out across an increasing time period of 5, 10, and 15 yr (Table 2), the t -test would allow to detect increasingly small differences in population abundances between the two landscape types (see Fig. 2D–F). The effect of increasing the sample size through adding further years was very pronounced in the first 5 yr and then

Table 2. Relative detection capacity (D_N) of Welch's t -test ($\alpha = 0.05$, $\beta = 0.2$) comparing mean butterfly abundance across two broad landscape types: landscapes with prevalence of agricultural habitats (Ag) or with low agricultural activity (non-Ag).

Timespan	Butterfly group/species	n	Dif.	SE	Sig	D_N (%)
5 yr	All species	445	-3.5	0.64	***	12.0
	Open habitat species	438	-3.8	0.96	***	15.8
	Mobile species	438	-0.7	0.49	0	12.9
	Grassland indicators	432	-1.5	0.68	*	14.2
	<i>Aricia cramera</i>	140	-0.7	1.19	0	28.0
	<i>Carcharodus alceae</i>	123	-3.0	0.82	***	24.4
	<i>Lasiommata megera</i>	366	-3.7	0.50	***	12.2
	<i>Lycaena phlaeas</i>	219	-3.5	0.65	***	16.7
	<i>Pararge aegeria</i>	315	-1.3	0.72	0	17.5
	<i>Pieris napi</i>	190	2.7	0.75	***	21.8
	<i>Polyommatus icarus</i>	320	-3.0	1.89	0	29.9
	<i>Thymelicus acteon</i>	152	-9.4	1.50	0	32.4
10 yr	All species	1028	-3.5	0.41	***	7.8
	Open habitat species	1004	-4.0	0.59	***	10.0
	Mobile species	1013	-0.7	0.34	*	8.8
	Grassland indicators	998	-1.4	0.44	**	9.4
	<i>Aricia cramera</i>	315	-1.0	0.86	0	20.3
	<i>Carcharodus alceae</i>	293	-2.7	0.60	***	17.7
	<i>Lasiommata megera</i>	824	-3.3	0.35	***	8.5
	<i>Lycaena phlaeas</i>	503	-3.4	0.43	***	11.0
	<i>Pararge aegeria</i>	706	-1.0	0.48	*	11.6
	<i>Pieris napi</i>	344	2.7	0.67	***	19.6
	<i>Polyommatus icarus</i>	728	-3.4	1.09	**	17.7
	<i>Thymelicus acteon</i>	378	-3.3	0.96	0	21.5
15 yr	All species	1540	-3.5	0.34	***	6.3
	Open habitat species	1501	-4.6	0.49	***	8.1
	Mobile species	1514	-1.0	0.27	***	6.8
	Grassland indicators	1489	-1.6	0.36	***	7.8
	<i>Aricia cramera</i>	453	-1.6	0.64	*	15.9
	<i>Carcharodus alceae</i>	412	-3.4	0.60	***	16.7
	<i>Lasiommata megera</i>	1230	-3.4	0.29	***	6.9
	<i>Lycaena phlaeas</i>	730	-3.3	0.36	***	9.2
	<i>Pararge aegeria</i>	1018	-1.1	0.42	*	10.4
	<i>Pieris napi</i>	730	2.8	0.60	***	17.0
	<i>Polyommatus icarus</i>	1086	-2.9	0.82	***	13.7
	<i>Thymelicus acteon</i>	540	-3.8	0.92	***	20.1

Notes: The analysis was repeated for an increasing timescale of 5 yr (1999–2003), 10 yr (1999–2008), and 15 yr (1999–2013). The mean sample size (n) was the mean number of sections in each landscape type, the difference in abundance (Dif.) is butterflies/km section of Ag compared to non-Ag sections, the SE of the t -test was calculated according to Eq. 1, and significant differences (Sig.) in Ag compared to the control (non-Ag) are shown by asterisks: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

progressively leveled off toward the end of the 15-yr period.

There were differences in abundance of all the multispecies indicators and single species between landscape types, with few exceptions (Table 2). With 5 yr of monitoring data, D_N values were halved for most butterfly indicators regarding the mean values of D_N obtained in the annual analysis (Fig. 2D–F) indicating a better

detection capacity. When means were compared using the entire 15-yr dataset, D_N dropped from a mean annual detection capacity of around 30% to a D_N below a 10% population change, 6.3% for all species, 8.1% for open habitat species, 6.8% for mobile species, and 7.8% for grassland indicators (Fig 2D). Regarding single species, the capacity of the t -test to detect changes in abundance between the two landscape types also

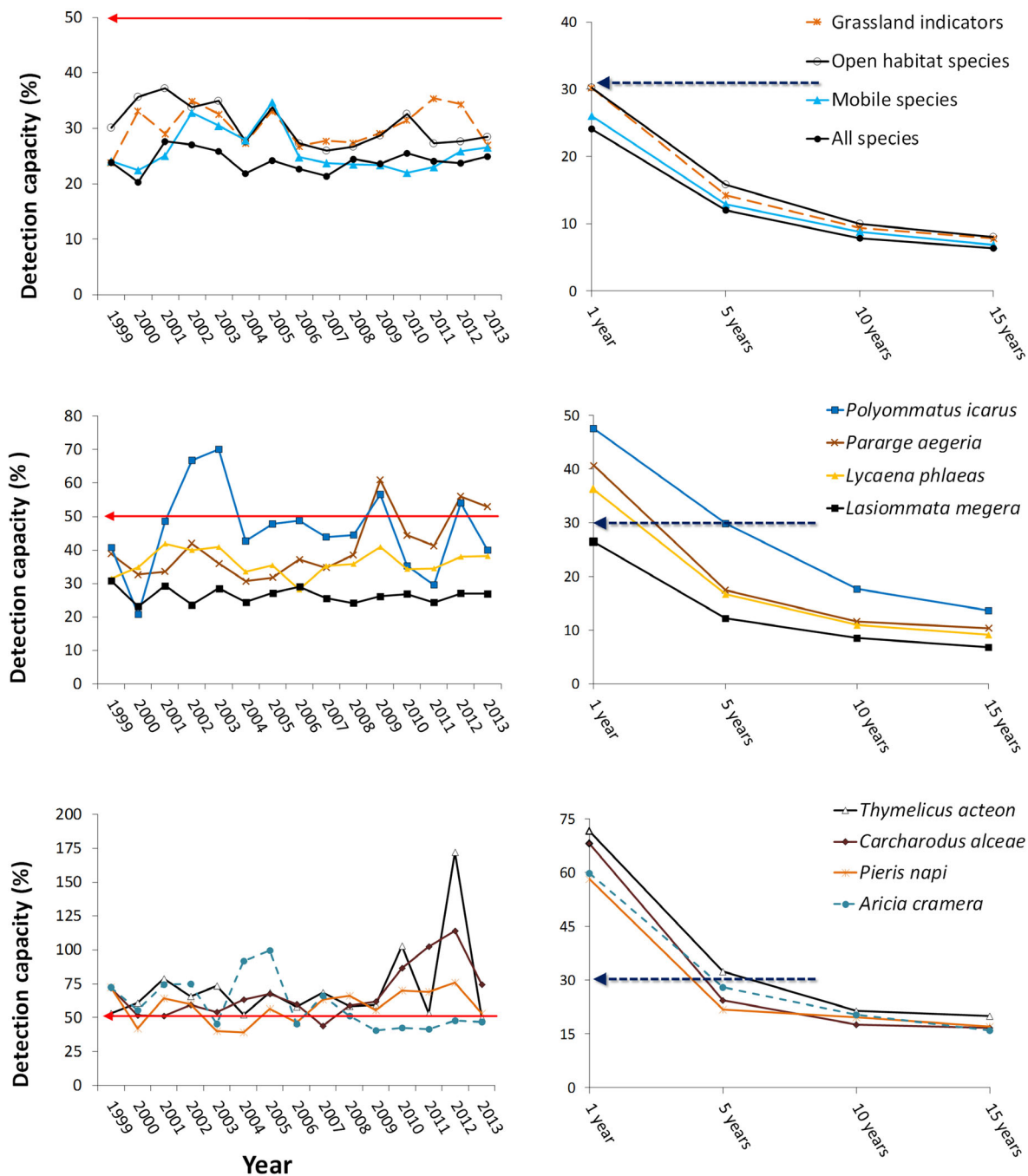


Fig. 2. The figures show the relative detection capacity (D_N) of Welch's t -test ($\alpha = 0.05$ and $\beta = 0.2$) to detect differences in butterfly abundance across two landscape use types (transects located in intensive agricultural areas compared to transects located in non-agricultural areas). When compared within any single year (A, B, C), the t -test would be capable of detecting population changes below 30% of the test population (non-agricultural areas) only when butterflies are aggregated in multispecies indicators (A); when abundance of single species was compared in the two transect types within single years, D_N was only below 30% for *Lasiommata megera* (B); for

(Fig. 2. Continued)

most widespread species, D_N was below 50% (represented by the red arrow), but as sample size and abundance decreased and variability increased, the t -test would be unable to detect population changes below 50% of the test population (C). As data from an increasing number of years were used (5, 10, or 15 yr) compared to a single year, D_N improved (a smaller population change could be detected). For the multispecies groups, in most cases two years would be sufficient to detect a 30% population change (D_N). Regarding single species, the number of years necessary for the t -test to detect a 30% population difference in abundance varied; for the seven widespread species, 5 yr would already allow to detect a 30% population change (E and F); however, in the case of the less abundant and less widespread *Thymelicus acteon*, at least 6 or 7 yr would be necessary for the t -test to have the capacity to detect a change below 30% (F). Note the different scales of the y -axis in the figures.

improved rapidly with the addition of further years; after five years of monitoring, it would be possible to detect a 30% change in abundance of all species except for *T. acteon*; after 15 yr, D_N values of single species (see Fig. 2E, F) were, on average, 13.7%, but there were considerable differences between species, ranging from 6.9% (*L. megera*) to 20.1% (*T. acteon*).

Factors influencing detection capacity

The magnitude of the test's SE greatly influenced both the eventual significance of the t -test and its ability to detect differences between population means (Eq. 2). The smaller the value of SE, the higher is the value of the contrast statistic, and therefore, the t -test is more likely to be significant. Consequently, when the value of SE is small, the test is capable of detecting very small differences between populations (low values of D), and this translates into a high detection capacity.

There was a fairly strong direct linear relationship ($R^2 = 0.75$) between the relative detection capacity (D_N) of the t -test and the SE of the contrast statistic, described by

$$D_N = 16.0 \times SE + 11.3 \quad (4)$$

In addition, the strong dependence ($R^2 = 0.80$) of detection capacity (D_N) on sample size (n) can be represented by an inverse potential relationship between the two variables:

$$D_N = 319.5n^{-0.51} \quad (5)$$

For instance, the species with the worst D_N values, *Aricia cramera*, *Carcharodus alceae*, and *T. acteon* (Figs. 2C, F) also had the highest SE values (see Tables 1, 2; Appendix S3). However, there were some exceptions; for instance, considering the annual analysis, *Polyommatus icarus* had

a relatively high variability but its high sample size (mean of 145 sections) and large difference between mean abundance of samples (2.76 individuals per km) resulted in an acceptable D_N ; *Pieris napi*, conversely, despite a lower SE (average of 1.95 butterflies per km), its low sample size (mean of 68 sections), and smaller difference in mean abundance (3.00 individuals per km) resulted in a poor D_N (average of 58.3%).

DISCUSSION

The results of this study indicate that data generated by the CBMS, used "as is," have a very high sensitivity to detect impacts of modified agricultural practices on butterfly populations, provided that a suitable indicator is chosen.

Detection capacity of the CBMS data

This study shows that a t -test, carried out on data from a well-established butterfly monitoring scheme, would be capable of detecting an acceptably small change in butterfly abundance between two transect types, here exemplified by transects in intensive agricultural landscapes (Ag) compared to areas with a lower agricultural activity (non-Ag). When the butterfly abundance in Ag transects was compared to the non-Ag transects, using data from the entire 15-yr period, the relative detection capacity (D_N) was below 30% for 52 of the 135 species in the dataset. The species that had a good detection capacity were generally those most frequent across the landscape (translated into a large sample size) and with a relatively low variability. Considering the 12 selected butterfly indicators, D_N was below 25% of the population abundance. This is a very good relative detection capacity, considering that population changes of 25–50% are considered

acceptable in field trials, for instance, to assess risks of GM crops on nontarget organisms (Duan et al. 2006, Lopez et al. 2005, Perry et al. 2003). Detection capacity was very good because CBMS samples were very large (>300 sections) and the standard errors relatively moderate (<0.95 individuals per km).

In contrast, when comparing mean butterfly abundance across the two transect types within the same year, instead of aggregating years, the values of D_N were generally much poorer (rarely below 30%) because samples were relatively small (<30 sections) and the standard errors were relatively large (>1.9 individuals per km). For instance, four of the eight single species selected for the study (*A. cramera*, *C. alceae*, *P. napi*, and *T. acteon*) may not be suitable for monitoring because the *t*-test would only be capable of detecting changes if the annual population decrease was over 50%. Conversely, it would be possible to detect a population change below 30% of the control population of the multispecies indicators all species and mobile species; and the single species *L. megera*, even using data from a single year. Detection capacity increased rapidly as further years were added due to a greater sample size. After 5 yr, D_N was below 30% for all indicators tested except *T. acteon*. Similarly, in a recent study in which butterflies were sampled to determine the effort needed to detect a 30% reduction in abundance due to GM maize cultivation, it was found that recording 9–25 transects during 3 yr would have sufficient statistical power (Lang et al. 2019).

In this study, detection capacity depended heavily on sample size. This was also reported by Lang and Bühler (2012) in two Swiss butterfly monitoring schemes when mean annual abundance of multispecies indicators or single species was analyzed; these authors calculated the sample size necessary for detecting changes in abundance of butterfly populations when pooled or single species' data were used for calculations. The values describing the relationship between sample size and the detection capacity were not very different to those found in the present work except for single species in which the detection capacity was more variable than in our analysis. As Lang and Bühler (2012) only disposed of two datasets for some sites, they were unable to test the detection capacity using longer time-series.

Worse results were obtained by Aviron et al. (2009) when using a dataset from a monitoring project on ecological compensation areas and biodiversity in Switzerland; this was mainly due to the low number of years (and therefore sample size) used for records. Aviron et al. (2009) consequently concluded that case-specific monitoring would not be appropriate for detecting possible effects of cultivation of GM *Bt* crops on butterflies because in order to detect an effect around 30%, over 100 pairs of fields would need to be sampled.

Regarding the butterfly indicators tested, the multispecies indicators, as expected (e.g., Brereton et al. 2011, Lang and Bühler 2012), performed much better than the single species but the drawback of using multispecies indicators is that results are difficult to interpret and effects may go unnoticed, that is, decreases of single species can often be masked by increases of others species. This is exemplified here by the fact that there were barely any differences in abundance of mobile species when tested annually despite significant differences for most of the single species. The most interesting species for monitoring agricultural impacts are those that are frequent across the landscape and have the lowest standard deviations, that is, those with less clumped distributions. Among these, there was *L. megera*, which is a common widespread butterfly in grassland (EEA 2013) whose populations are declining across northwestern Europe, possibly due to climate change (Van Dyck et al. 2015). Another interesting indicator species in the study area was *Pieris napi*, which, despite a poor detection capacity, was the only species more abundant in farmland. This species does not usually feed on crop plants (García-Barros et al. 2013). In the arid Mediterranean climate, its presence in agricultural areas has been explained by the availability of humid environments that help to buffer the effects of extreme temperatures and droughts (Carnicer et al. 2019), which are increasingly aggravated by global climate change. Finally, *T. acteon*, with its status as a Red List Species (Van Swaay et al. 2010), had a rather poor relative detection capacity. This is often the case for endangered organisms, although the need to protect them from harm is very high, they are not frequent enough to allow for suitable statistical analysis.

A case study: the CBMS for general surveillance of GM crops

There are many practical cases in which it is necessary to obtain reliable information on the effects of changes in agricultural practice on the receiving environment. For instance, in the case of genetically modified (GM) crops in the EU, post-market environmental monitoring (PMEM) is compulsory (Directive 2001/18/EC), and EU regulations require the implementation of general surveillance (GS) plans for long-term monitoring (EFSA 2010). Much effort has been devoted in Europe to generate data for PMEM of GMOs, for butterflies (i.e., Lang et al. 2019), and other taxa, particularly in the UK (Clark et al. 2006) and Spain (De la Poza et al. 2005, Comas et al. 2014). In Spain, GM corn has been grown in thousands of hectares in the last 15 yr (ISAAA 2016), but GS has yet to be fully implemented. In order to reduce costs and increase monitoring practicability, regulation authorities recommend companies to use ESNs (EFSA 2011) rather than implementing field studies. There are contrasting opinions on the utility of these networks for GS of GM crops; whereas Smets et al. (2014) consider that these networks would only provide information on the baseline variation of indicators, other studies such as Lang and Bühler (2012) and the present study indicate that the data obtained by butterfly monitoring schemes are sensitive enough to detect changes in populations of the indicator organisms.

Whereas field studies for environmental risk assessment (ERA) of GM crops (or plant protection products) are designed specifically to detect effects, ESNs aim to obtain more general information on population dynamics. This results in differences in the factors influencing the capacity of the *t*-test to detect population change. For instance, when authors analyzed data from over 12 yr of field trials for ERA of GM corn, using several arthropod taxa to detect eventual changes in nontarget arthropod abundance due to cultivation of *Bt* corn; taxon abundance was the most influential factor determining the relative detection capacity (D_N) of Welch's *t*-test (Albajes et al. 2013). The D_N of a taxon improved as its abundance increased because its relative variability decreased. Since the number of replicates in the experimental trials was fairly constant (3–4 blocks), the SE of the contrast statistic

was determined mainly by the variability of the sample, not by sample size as in this study.

This study shows that data from the CBMS could be used to monitor post-market effects of GM crops on butterflies, given its high capacity to detect possible effects. This is a relevant point in GS of GM crops because environmental impacts of GM crops in the field, if any, would appear to be of low magnitude (Naranjo 2009, Romeis et al. 2009, Pleasants and Oberhauser 2012, Comas et al. 2014), and therefore, a high detection capacity of statistical tests is needed.

Nevertheless, a drawback of the CBMS, compared to field trials, is that it mainly samples seminatural habitats so there may be years when few transects are located in the vicinity of GM, at the beginning of *Bt* corn deployment, and non-GM cornfields, when *Bt* corn has been successfully established in the area. This would result in a reduction of detection capacity. To address this issue, EFSA (2014) recommended the possible increase of transect number, which would also increase costs (Schmeller and Henle 2008). Nevertheless, this problem should be mitigated in future because butterfly monitoring schemes are attempting to increase number and spatial uniformity of monitoring sites in agricultural landscapes, where butterfly populations may suffer more pressures from agricultural practices (Brereton et al. 2010). Additionally, the increase in the number of sites could also be partially compensated by a lower sampling frequency (Brereton et al. 2010). Notwithstanding the pitfalls of ESN data, it is often not possible to use field trials due to the costs involved, so data from these monitoring schemes represent the only practicable option for environmental monitoring.

Conclusion: Is the CBMS suitable for monitoring?

With probability levels fixed at $\alpha = 0.05$ and $\beta = 0.2$, a Welch's *t*-test on CBMS data had the capacity to detect changes in abundance between 6% and 20% of the selected butterfly indicators when samples were compared across a 15-yr period. Detection capacity was found to depend mainly on sample size; thus, species that were more frequent across the landscape tended to have higher detection capacities. When mean butterfly abundances between transects were compared within the same year, the sensitivity of the *t*-test was much lower but

it would still be possible to detect population changes between 24% and 50% in eight out of the 12 butterfly indicators tested. Detection capacity rapidly improved with the addition of further years as this greatly increased sample size, and after 5 yr of monitoring, it would be possible to detect differences in abundance below a 30% threshold. In conclusion, this study shows that, despite some pitfalls, data from existing environmental survey networks do have the potential to be used for environmental monitoring of agrienvironmental measures to inform shareholders and policymakers. In the specific case of GM crops, the CBMS could be used to monitor post-market effects on butterflies, given its high capacity to detect possible effects. However, the specific testing approach would have to be adapted to the nature of the expected agricultural impacts to be monitored and to the particular characteristics and limitations of the monitoring scheme data.

ACKNOWLEDGMENTS

We deeply thank the CBMS volunteers and collaborators. We also thank two anonymous reviewers for their comments. This work was partially supported by the Spanish Government-funded project AGL2011-23996 and a FI-DGR scholarship to M. Lee from the Catalan Government. M. S. Lee wrote the manuscript and analyzed the data. J. Comas is a statistics consultant, analyzed the data, and wrote the manuscript. R. Albajes is the director of the study, analyzed the data, and wrote the manuscript. C. Stefanescu provided data from CBMS, wrote the manuscript, is a butterfly expert.

LITERATURE CITED

- Albajes, R., B. Lumbierres, and X. Pons. 2013. Representative taxa in field trials for environmental risk assessment of genetically modified maize. *Bulletin of Entomological Research* 103:724–733.
- Aviron, S., O. Sanvido, J. Romeis, F. Herzog, and F. Bigler. 2009. Case-specific monitoring of butterflies to determine potential effects of transgenic *Bt*-maize in Switzerland. *Agriculture, Ecosystems and Environment* 131:137–144.
- Brereton, T. M., K. L. Cruickshanks, K. Risely, D. G. Noble, and D. B. Roy. 2010. Developing and launching a wider countryside butterfly survey across the United Kingdom. *Journal of Insect Conservation* 15:279–290.
- Brereton, T., D. B. Roy, I. Middlebrook, M. Botham, and M. Warren. 2011. The development of butterfly indicators in the United Kingdom and assessments in 2010. *Journal of Insect Conservation* 15:139–151.
- Carnicer, J., C. Stefanescu, M. Vives-Inglà, C. López, S. Cortizas, C. Wheat, R. Vila, J. Llusà, and J. Peñuelas. 2019. Phenotypic biomarkers of climatic impacts on declining insect populations: a key role for decadal drought, thermal buffering and amplification effects and host plant dynamics. *Journal of Animal Ecology* 88:376–391.
- CBMS. 2016. Catalan butterfly monitoring scheme. Museu d'Història Natural de Granollers, Granollers, Spain. www.catalanbms.org
- Clark, S. J., P. Rothery, and J. N. Perry. 2006. Farm Scale Evaluations of spring-sown genetically modified herbicide-tolerant crops: a statistical assessment. *Proceedings of the Royal Society. Biological Sciences* 273:237–243.
- Comas, J., B. Lumbierres, X. Pons, and R. Albajes. 2013. Ex-ante determination of the capacity of field tests to detect effects of genetically modified corn on nontarget arthropods. *Journal of Economic Entomology* 106:1659–1668.
- Comas, C., B. Lumbierres, X. Pons, and R. Albajes. 2014. No effects of *Bacillus thuringiensis* maize on nontarget organisms in the field in southern Europe: a meta-analysis of 26 arthropod taxa. *Transgenic Research* 23:135–143.
- de la Poza, M., X. Pons, G. P. Farinós, C. López, F. Ortego, M. Eizaguirre, P. Castañera, and R. Albajes. 2005. Impact of farm-scale *Bt* maize on abundance of predatory arthropods in Spain. *Crop Protection* 24:677–684.
- Di Stefano, J. 2003. How much power is enough? Against the development of an arbitrary convention for statistical power calculations. *Functional Ecology* 17:707–709.
- Duan, J. J., C. Jiang, G. P. Head, M. A. Bhatti, D. P. Ward, S. L. Levine, T. E. Nickson, and M. A. Nemeth. 2006. Statistical power analysis of a 2-year field study and design of experiments to evaluate non-target effects of genetically modified *Bacillus thuringiensis* corn. *Ecological Entomology* 31:521–531.
- EEA, European Environment Agency. 2013. The European Grassland Butterfly Indicator: 1990-2011, Technical report No 11/2013. Publications Office of the European Union, Luxembourg, Luxembourg.
- EFSA Panel on Genetically Modified Organisms. 2010. Guidance on the Environmental Risk Assessment of genetically modified plants. *EFSA Journal* 10:1877.
- EFSA Panel on Genetically Modified Organisms. 2011. Guidance on the Post-Market Environmental

- Monitoring. PMEM. of genetically modified plants. EFSA Journal 9:2316.
- EFSA Panel on Genetically Modified Organisms. 2014. Scientific Opinion on the use of existing environmental surveillance networks to support the post-market environmental monitoring of genetically modified plants. EFSA Journal 12:3883.
- Elzinga, C. L., D. W. Salzer, J. W. Willoughby, and J. P. Gibbs. 2001. Monitoring plant and animal populations. Blackwell Science, New York, New York, USA.
- ESRI. 2008. ArcGIS Desktop 9.3. Environmental Systems Research Institute, Redlands, California, USA.
- Fagerland, M. W. 2012. *t*-tests, non-parametric tests, and large studies—a paradox of statistical practice? BMC Medical Research Methodology 12:78.
- Feber, R. E., P. J. Johnson, L. G. Firbank, A. Hopkins, and D. W. Macdonald. 2007. A comparison of butterfly populations on organically and conventionally managed farmland. Journal of Zoology 273:30–39.
- Field, S. A., P. J. O'Connor, A. J. Tyre, and H. P. Possingham. 2007. Making monitoring meaningful. Austral Ecology 32:485–491.
- García-Barros, E., M. L. Munguira, C. Stefanescu, and A. Vives Moreno. 2013. Lepidoptera: Papilionoidea. Fauna Ibérica. Volume 37. CSIC, Madrid, Spain.
- Geijzendorffer, I. R., and P. K. Roche. 2013. Can biodiversity monitoring schemes provide indicators for ecosystem services? Ecological Indicators 33:148–157.
- Hails, R. S., S. Freeman, P. Henrys, and G. Siriwardena. 2012. Determining and increasing the sensitivity of existing environmental surveillance monitoring networks to detect unanticipated effects that may occur in the environment in response to the cultivation of genetically modified crops. SFFS0001: final Report. DEFRA, London, UK.
- Hazewinkel, M. 2002. Encyclopaedia of mathematics. Springer-Verlag, Berlin, Germany.
- Herrando, S., L. Brotons, M. Anton, F. Páramo, D. Villeró, J. Quesada, and C. Stefanescu. 2016. Assessing impacts of land abandonment on Mediterranean biodiversity using indicators based on bird and butterfly monitoring data. Environmental Conservation 43:69–78.
- ICGC. 2013. Institut Cartogràfic i Geològic de Catalunya. Catalan Institution of Cartography and Geology, Catalunya, Spain. www.icgc.cat
- ISAAA. 2016. Global Status of Commercialized Biotech/GM Crops: 2016. ISAAA Brief, Ithaca, New York, USA.
- Lang, A. 2004. Monitoring the impact of *Bt* maize on butterflies in the field: estimation of required sample sizes. Environmental Biosafety Research 3:55–66.
- Lang, A., and C. Bühler. 2012. Estimation of required sampling effort for monitoring the possible effects of transgenic crops on butterflies: lessons from long-term monitoring schemes in Switzerland. Ecological Indicators 13:29–36.
- Lang, A., F. Kallhardt, M. Lee, J. Loos, M. Molander, I. Muntean, L. Pettersson, L. Rákósy, C. Stefanescu, and A. Messéan. 2019. Monitoring environmental effects on farmland Lepidoptera: Does necessary sampling effort vary between different bio-geographic regions in Europe? Ecological Indicators 102:791–800.
- Lee, M. S., and R. Albajes. 2013. Butterflies for post market environmental monitoring of GM maize in Spain. IOBC/WPRS Bulletin 97:63–72.
- Lopez, M. D., J. R. Prasifka, D. J. Bruck, and L. C. Lewis. 2005. Utility of ground beetle species in field tests of potential nontarget effects of *Bt* crops. Environmental Entomology 34:1317–1324.
- Morecroft, M. D., et al. 2009. The UK Environmental Change Network: emerging trends in the composition of plant and animal communities and the physical environment. Biological Conservation 142:2814–2832.
- Naranjo, S. 2009. Impacts of *Bt* crops on non-target invertebrates and insecticide use patterns. CAB Reviews: perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources 4.
- Perry, J., P. Rothery, and S. Clark. 2003. Design, analysis and statistical power of the Farm-Scale Evaluations of genetically modified herbicide-tolerant crops. Journal of Applied Ecology 40:17–31.
- Pleasants, J. M., and K. S. Oberhauser. 2012. Milkweed loss in agricultural fields because of herbicide use: effect on the monarch butterfly population. Insect Conservation and Diversity 6:135–144.
- Pollard, E., and T. J. Yates. 1993. Monitoring butterflies for ecology and conservation: the British butterfly monitoring scheme. Chapman and Hall, London, UK.
- R Development Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ripple, W. J., C. Wolf, T. M. Newsome, M. Galetti, M. Alamgir, E. Crist, M. I. Mahmoud, and W. F. Laurance. 2017. World scientists' warning to humanity: a second notice. BioScience 67:1026–1028.
- Romeis, J., M. Meissle, A. Raybould, and R. L. Hellmich. 2009. Impact of insect-resistant transgenic crops on aboveground non-target arthropods. Pages 165–198 in N. Ferry and A. M. R.

- Gatehouse, editors. Environmental impact of genetically modified crops. CABI Publishing, Wallingford, UK.
- Schmeller, D. S., and K. Henle. 2008. Cultivation of genetically modified organisms: resource needs for monitoring adverse effects on biodiversity. *Biodiversity and Conservation* 17:3551–3558.
- Schmucki, R., G. Pe'er, D. B. Roy, C. Stefanescu, C. A. M. Van Swaay, T. H. Oliver, and R. Julliard. 2016. A regionally informed abundance index for supporting integrative analyses across butterfly monitoring schemes. *Journal of Applied Ecology* 53:501–510.
- Smets, G., et al. 2014. The use of existing environmental networks for the post-market monitoring of GM crop cultivation in the EU. *Environmental Science: processes and Impacts* 16:1754.
- Stefanescu, C., J. Carnicer, and J. Peñuelas. 2011a. Determinants of species richness in generalist and specialist Mediterranean butterflies: the negative synergistic forces of climate and habitat change. *Ecography* 34:353–363.
- Stefanescu, C., I. Torre, J. Jubany, and F. Páramo. 2011b. Recent trends in butterfly populations from north-east Spain and Andorra in the light of habitat and climate change. *Journal of Insect Conservation* 15:83–93.
- Thomas, J. A. 2005. Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. *Philosophical Transactions of the Royal Society of London B* 360:339–357.
- Tscharntke, T., A. M. Klein, A. Kruess, I. Steffan-Dewenter, and C. Thies. 2005. Landscape perspectives on agricultural intensification and biodiversity - ecosystem service management. *Ecology Letters* 8:857–874.
- Van Dyck, H., D. Bonte, R. Puls, K. Gotthard, and D. Maes. 2015. The lost generation hypothesis: Could climate change drive ectotherms into a developmental trap? *Oikos* 124:54–61.
- Van Swaay, C. A. M., P. Nowicki, J. Settele, and A. J. Van Strien. 2008. Butterfly monitoring in Europe: methods, applications and perspectives. *Biodiversity and Conservation* 17:3455–3469.
- Van Swaay, C., et al. 2010. European red list of butterflies. Publications Office of the European Union, Luxembourg, Luxembourg.
- Welch, B. L. 1947. The generalisation of "Student's" problem when several different population variances are involved. *Biometrika* 34:28–35.

SUPPORTING INFORMATION

Additional Supporting Information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/ecs2.3004/full>